A Heuristic Position-Based Leakage and Dynamic Power Reduction Technique in the Kogge-Stone Adder

Elaheh Sadredini, Luonan Wang
ECE 6332, University of Virginia
{es9bt,lw3mz}@virginia.edu

ABSTRACT

Power consumption is an increasingly serious problem in very large scale integrated circuit. Furthermore, adders are important component in DSPs and microprocessors. In this research, first we investigate power-delay trade-off for Carry Look-ahead Adder (CLA) and Kogge-Stone adder (KSA). We use header and footer sleep transistors as power gating techniques and change their widths. Header sleep transistor is better than footer regarding leakage reduction and propagation delay. Second, we focus on KSA and propose a position-based technique to reduce leakage and dynamic power without posing large delay penalty. The proposed approach is compared with the basic power gating techniques and results in 24% leakage power and 52% dynamic power reduction with only 10% delay penalty. We use freePDK technology for our simulations.

Key-words; carry look-ahead adder; kogge-stone adder, power reduction, propagation dealy.

1. INTRODUCTION

One of the challenge with technology scaling is the rapid increase in sub-threshold leakage power due to reduction in V_{th}. If the present scaling trend continue, it is expected that the sub-threshold leakage power will become a considerable constituent of the total dissipated power. Especially in the portable systems such as smart phones and tablets, battery life span is of great concern. When the device is in standby mode, it still consumes energy through leakage current from the cells which are cut-off. Although the leakage current is less than the active current, it will gradually consume most of the energy through a long period of time and long standby time in such devices are common [5]. In such a system it becomes crucial to identify techniques to reduce this leakage power component. Previous work has been done to use various power reduction technique to limit the leakage current. In [Multi-Mode paper], a single PMOS is added in parallel with NMOS at the footer which minimizes the ground bound noise induced by power mode transitions of the power gating structures yielding an expanded space for power-performance tradeoffs that supports three different power modes, RUN, PARK and COLD. In [7], "sleepy stack" approach is presented which forces a stack effect by breaking down an existing transistor into two half size transistors. The forced stack approach can achieve huge leakage power saving while retaining the logic state [7].

Full adders are important components in applications such as digital signal processors and microprocessors. Apart from the basic addition, adders are also used in performing useful operations such as subtraction, multiplication, division and address calculation. In most of these systems the adder lies in the critical path that determines the overall performance of the system. Furthermore, energy reduction of functional units is a very important concern for high-end superscalar processors.

Implementations of different adder cells are proved to be efficient for reduction in leakage power and propagation delay. There are always tradeoff between power and delay.

This research has two parts. First, we investigate power-delay trade-off for carry look-ahead adder and Kogge-Stone adder with two power gating technique and try to figure out which technique is better form delay and leakage aspect for these adders. In the second part, we focus on Kogge-Stone adder and pose some research questions on how to reduce leakage power and dynamic power without significant penalty in delay. We also talk about input vector control and its effect on leakage power.

In this project we choose coarse-grained power-gating design instead of fine-grained power-gating. The major concern is still the area. The main advantage of using fine-grained power-gating technique is that the virtual power source net are short and it is easy to build multiple layer of schematics from the existed cells. The disadvantage for this technique is easy to observe, the overall area will be significantly large due to attaching a power gate to every single cell. The coarse-grained power-gating technique has less area incurred and less sensitivity to PVT variation. Thus in the following experiment, we'll use coarse-grained technique for power gating.

2. CARRY-LOOKAHEAD ADDER

Carry Look-ahead Adder (CLA) improves speed by reducing the amount of time required to determine carry bits. Kogge-Stone adder is parallel prefix form of CLA which is much faster. In this part, we build an 8-bit CLA from two 4-bit CLA and 8-bit Kogge-Stone adder (KSA). Then we measure the power and delay of the CLA and KSA with two power gating techniques: only footer and only header. The research question here is which technique, header or footer has the better performance and leakage reduction on CLA and KSA.

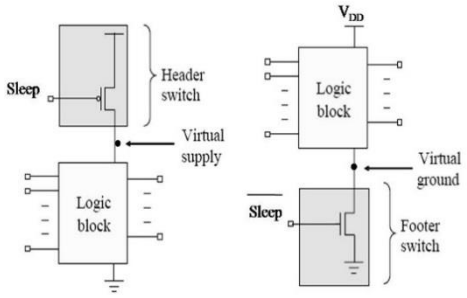

Figure [1]. Power gating with header and footer implementations

$$I_D = I_0 {}^{W}/_{L}\, 10^{\frac{V_{GS} - V_T + \eta V_{DS}}{s}} \quad (1)$$

According to the equation for leakage current, the threshold

voltage has an exponential influence on the final leakage current; thus using transistor with high V_{th} can significantly reduce the leakage current. Using high-VT CMOS to cut-off logic block to either VDD or GND will reduce a large amount of the leakage current. We add NMOS as the footer between the adders and the ground and use ocean script to sweep the width of the footer from 50nm to 500nm with 50nm step size. The same measurement is done for adding PMOS as the header between the adders and VDD. From Fig. 2 to Fig. 5, we can observe that using only header causes more reduction in the leakage rather than footer. In addition, propagation delay penalty is less, when using header as a sleep transistor instead of footer.

We can also see that the propagation delay for the KSA are less than those in CLA. NMOS tends to be leakier than PMOS due to NBTI (Negative-Bias Temperature Instability) effect. The NBTI effect increases V_{th} over time and makes PMOS transistor less leaky. However, NMOS transistor has the advantage of having less area compared to PMOS and more driving capability than PMOS [2].

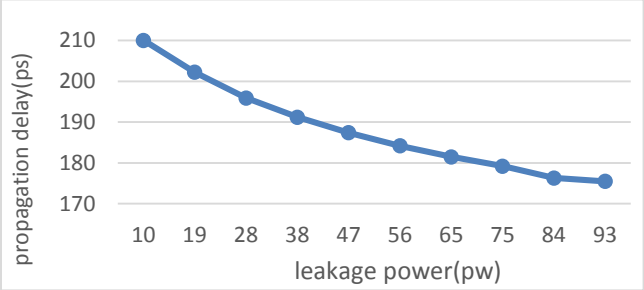

Figure [2]. Power-delay curve using only footer for power gating on 8-bit CLA

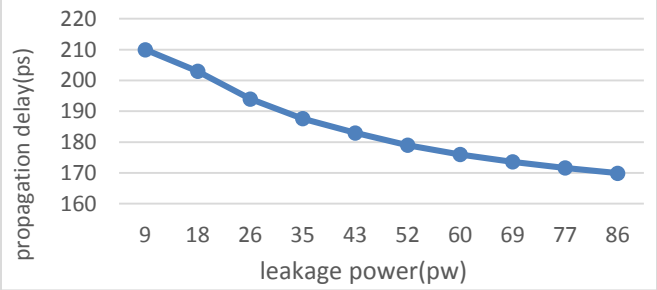

Figure [3]. Power-delay curve using only header for power gating on 8-bit CLA

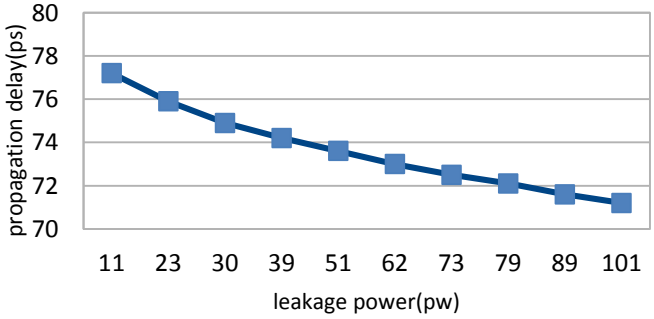

Figure [4]. Power-delay curve using only footer for power gating on 8-bit KSA

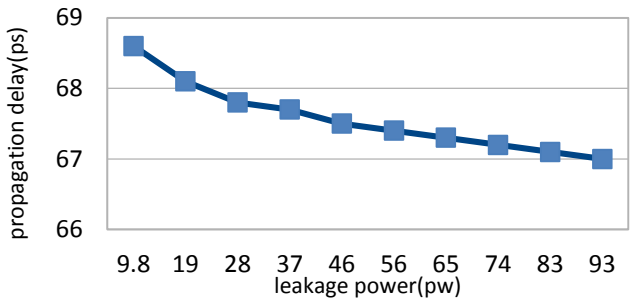

Figure [5]. Power-delay curve using only header for power gating on 8-bit KSA

3. KOGGE-STONE ADDER

Kogge-stone adder (KSA) is parallel prefix form of carry look-ahead adder. It is highly considered as the fastest adder design [4] and use widely used in high-performance processors. It generates the carry signal in (log n) time, where n is the number of bits per input. KSA is consisted of three main blocks; 1. PG generation, 2. Dot blocks, 3. Sum generation. Fig. [6] shows an example of a 4 bit Kogge-stone adder with zero carry in.

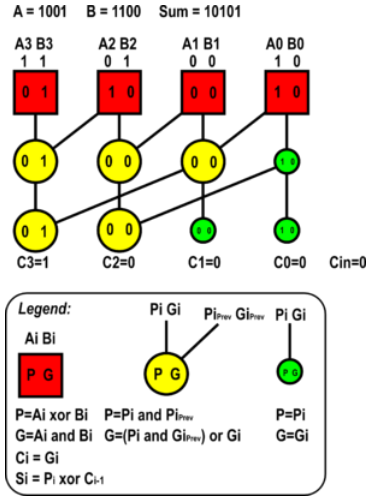

Figure [6]. Example of a 4-bit kogge-stone adder with zero carry-in [1]

First, we drew the schematic for the 8 bit and 16 bit KSA and then verified them. The load to be driven at the output is 1fF for each output sum bit. For the worst case delay, we looked for the input vector that activate critical path.

In the next parts, we want to pose some research questions and then try to answer them one by one.

3.1 Minimizing leakage power in KSA with IVC

The input vector control (IVC) problem is to find an input vector that minimize a circuit's leakage current. It takes advantage of the transistor stacking effect within the logic gates. The question here is what input vector for the KSA results in least leakage power in the standby mode? Exhaustive search using ocean takes too much time especially for the larger adders. From the KSA structure, dot blocks are dominant blocks; so we can find out the input vector which causes the least leakage dissipation for a dot block and

then, use vector forcing. Table 1 represents leakage power calculation for all possible input vectors in a dot block.

Table 1. Leakage power for all different inputs in dot block

P1	P2	G1	G2	Leakage (nw)
0	0	0	0	41.51 (I2)
0	0	0	1	44.02
0	0	1	0	55.53
0	0	1	1	42.77
0	0	0	0	60.04
0	0	0	1	51.78
0	1	1	0	50.66
0	1	1	1	39.77 (I1)
1	0	0	0	51.02
1	0	0	1	48.53
1	0	1	0	60.04
1	0	1	1	47.28
1	1	0	0	70.72 (I3)
1	1	0	1	62.46
1	1	1	0	61.34
1	1	1	1	50.45

I1 results in the lowest leakage power dissipation for the dot block. The problem here is that by using I1 as the input of the dot blocks, we cannot apply vector forcing to set the primary inputs of the circuit. Fortunately, I2, which is the next best input vector, is all zero and vector forcing can be easily applied. As a result by setting all inputs to zero, the lowest leakage power for KSA can be achieved. Table 2 proof this assertion.

Table 2. Leakage power for KSA

Input Vector	Leakage (uw)	
	8 bit KSA	16 bit KSA
All '0'	1.89	4.48
All '1'	1.96	4.61
1010 … 1010	1.96	5.97

Therefore, when the circuit is in the standby mode, by using scan chain and applying all zero inputs to the primary inputs of the circuit, minimum leakage dissipation will be achieved.

3.2 Where is the best place for applying power gating techniques?

In the common power gating designs, a single sleep transistor in the header or in the footer is used as a power gating technique to reduce leakage power. Imagine that because of some limitations like delay, we want to apply power gating structure partially to the circuit. Now, we want to decide where is the best place to add power gating structures in the KSA. In the other words, where we can add sleep transistors to benefit more and loos less? There are two approaches we can take. First, adding power gating to the parts of the circuit that does not increase critical path. Second, adding power gating structure to the parts of the circuit that are more potential on leaking more. As we mentioned earlier, dot blocks are dominant part of the circuit. Considering Table 2, I3 results in the worst leakage dissipation. Now the question is which level of the dot blocks has the most probability of getting I3 as their inputs. For realizing that, we need to apply all possible combination of primary inputs to the circuit. Considering that ocean is too slow to the simulation, we have written a program in SystemC to calculate the probability of different inputs in different level of dot blocks. We chose SystemC to take advantage of its timing diagram. Fig. 7 shows the number of having I3 in the each level dot blocks.

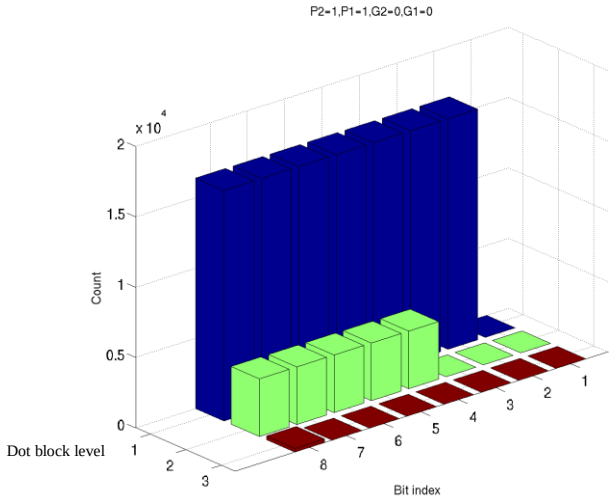

Figure [7]. Number of having I3 for each dot block in the 8-bit KSA

We can conclude from the results in the Fig. 7 that the first level dot block are more potential in getting I3 as their inputs. As a results, first level dot blocks are leaking more.

In addition, half of the sum blocks are not in the critical path and we can add power gating techniques there without concerning about the delay. For example in the 8-bit KSA, S0 to S3 get ready sooner than S4 to S7. In addition, we can decide that how much sleep transistor can be small to reduce the power as much as it can and not increasing worst case delay.

3.2.1 Adding power gating techniques in the non-critical path

As it is discussed above, half of the sum blocks in the KSA are not located in the critical path. So we can apply power gating techniques to those blocks and see their effects on the leakage power and active power.

Table 3. Power and delay results for adding power gating in non-critical path

Power Gating Technique	Leakage in Standby (uw)	Leakage in active (uw)	Dyn. Power (uw)	Delay (ps)	Leakage Imp (%)	Dyn. Imp (%)
Nothing	2.465	2.465	796	44	NA	NA
PMOS in header	2.189	2.465	560	44	11	29
NMOS in header	2.189	2.266	517	44	11	35

As the results shown, when we use NMOS or PMOS in the header, leakage power in the standby mode decreases. Also,

Table 4. Power and delay results for the proposed and common approaches

Power Gating Technique	Leakage in Standby (pw)	Leakage in active (uw)	Dyn. Power (uw)	Delay (ps)	Leakage Imp-Stand-by (%)	Leakage Imp-Active (%)	Dyn. Imp (%)	Delay increment (%)
Common approach	108.53	2.428	96.14	53.8	NA	NA	NA	NA
Proposed approach	82.248	1.873	45.51	60	24	22	52	10

dynamic power for both decreases because of the stacking effect. Interestingly, reduction in the dynamic power for NMOS is more than PMOS, because NMOS is not a good pass transistor for one logic. Therefore, the bias voltage reduces and causes reduction in the dynamic power. In addition, leakage power in the active mode has reduced for NMOS.

For calculating leakage power, all inputs are set to zero. PMOS and NMOS width are considered 180 nm and 90 nm respectively. In addition, for calculating dynamic power, a transition of inputs which makes a lot of changes in the circuit and internal signal have been considered.

3.2.2 Applying NMOS and PMOS sleep transistors in different parts of KSA

We assume that all the power gating techniques are applied to the header. For the sleep transistor, we can use NMOS or PMOS in the header as the sleep transistor. As Table 3 represents, leakage in the active mode decrease more, when using NMOS sleep transistor in the header instead of PMOS sleep transistor. Of course it is because of decrement the in virtual VDD. In addition, leakage in the active mode decreases. On the other hand, delay will increase (for the gates which are in the critical path). The question is in the KSA, in which parts is better to use header PMOS sleep transistor and where is better to use header NMOS sleep transistor?

In part 3-2, we saw that first level dot blocks have higher potential in leaking both in standby and active mode. Although using NMOS in header increase delay more than using PMOS in the header, adding NMOS sleep transistor in the first level save more power at the same delay penalty when header NMOS in the other levels. That's because the number of dot blocks are more in the first level and we can save more energy. Furthermore, Table 3 proved that using NMOS sleep transistor in the header save more power.

Fig. 8 represents our approach to spread NMOS and PMOS transistors in the KSA. For the green blocks, we apply header NMOS sleep transistor and for the orange blocks, we use header PMOS sleep transistor. We have demonstrated why it is worthy to use NMOS sleep transistor in the header in first level dot blocks and half of the sum generation blocks. Table 4 shows the leakage power in the standby mode, leakage power in the active mode, dynamic power, and delay for the proposed approach. We also compared all these results with the common technique which just uses one PMOS sleep transistor in the header for the whole circuit.

For the proposed approach, width of the NMOS sleep transistor is 90 nm and PMOS sleep transistor is 180 nm. For the common approach, the size of PMOS sleep transistor is considered 360 nm. So they are comparing in the fair situation.

The baseline for the comparison of the proposed approach is common approach. Form the results in the Table 4, we could reduce leakage and dynamic power significantly just by imposing 10% penalty in the delay.

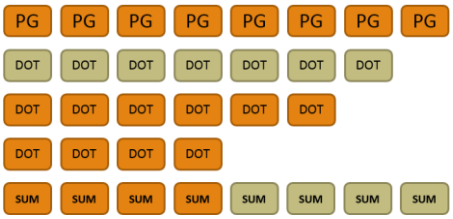

Figure [8]. Green blocks get NMOS header sleep transistor and Orange blocks get PMOS header sleep transistor

4. Conclusion

In this research, we calculated delay and power for different width in header and footer sleep transistors in CLA and KSA. For the KSA, we found the best input vector which results in least leakage power. Then we investigated the effect of PMOS and NMOS both in the header and then, we explore the best place to put header PMOS and header NMOS sleep transistor in the KSA. Simulation results show with only 10% delay increment, 52% dynamic power reduction and 24% leakage power reduction have been achieved. For the future works, we want to find the width threshold for the sleep transistor which are in the critical path. We also want to see the effect of proposed approach for the larger designs.

5. ACKNOWLEDGMENTS

We would like to thank Professor Ben Calhoun for his valuable advices and giving us the opportunity to working through this project and getting a lot of experience with the tools and technologies. We also wish to thank Farah, graduate teaching assistant for her guidance and expertise throughout this project.

6. REFERENCES

[1] http://en.wikipedia.org/wiki/Kogge%E2%80%93Stone_adder

[2] Kaijian Shi, David Howard. Sleep Transistor Design and Implementation – Simple Concepts Yet Challenges To Be Optimum. *International Symposium on VLSI Design, Automation and Test*, Apr. 2006

[3] Haihua Su, Frank Liu, Anirudh Devgan. Full Chip Leakage Estimation Considering Power Supply and Temperature Variations. *ISLPED'03* August 25-27, 2003

[4] S. Saddam Hussain, S, Mahaboob Basha, Comparatibve Design of 16-bit Sparse-Tree Rsfq Adder, *Journal of Engineering Research and Applications*, Vol. 4, Issue 9, June 2014, pp. 184-188.

[5] Amit Bakshi, Tribikram Pradhan. Implementation of Power Gatig Technique in CMOS Full Adder Cell to Reduce Leakage Power and Ground Bounce Noise for Mobile Application. *International Journal of Advanced Research in Electronics and Communication Engineering, Vol 1, Issue 4*, Oct 2012

[6] Suhwan Kim, Seoul Nat. Univ., Seoul, Kosonocky, S.V., Knebel, D.R., Stawiasz, K., A Multi-Mode Power Gating Structure for Low-Voltage Deep-Submicron CMOS ICs, *IEEE Transactions, IEEE Circuits and Systems Society*, Volume:54 , Issue: 7, July 2007.

[7] Jun Cheol Park, Vincent J. Mooney III, and Philipp Pfeifenberger. Sleepy Stack Reduction of Leakage Power,*PATMOS, LNCS 3254, pp 148-158, 2004.*